Introduction to Biostatistics

Dr Inimary Toby

Department of Biology

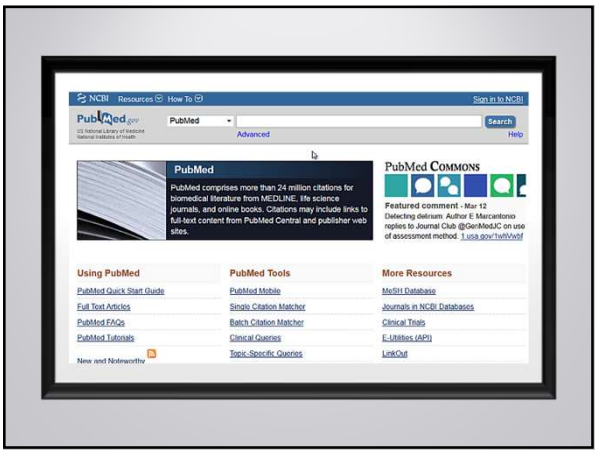
What to expect

- Review of the syllabus and course objectives:
  - Hands-on assignments
  - Case studies
  - Final projects
  - Overall course expectations

Why do we need Biostatistics?

Data is utilized and summarized frequently in research literature:

- Formal findings or original research work
- Data is readily open and available with "open access", FAIR standards, and other efforts to improve data transparency
- Examples can be viewed online at pubmed.nih.gov



From Archives of Surgery article, August 2000: -

- "Hypothesis: Surgeon-directed institutional peer review, associated with positive physician feedback, can decrease the morbidity and mortality rates associated with carotid endarterectomy."
- "Results: Stroke rate decreased from 3.8% (1993-1994) to 0% (1997-1998). The mortality rate decreased from 2.8% (1993-1994) to 0% (1997-1998). (Average) length of stay decreased from 4.7 days (1993-1994) to 2.6 days (1997-1998). The (average) total cost decreased from $13,344 (1993-1994) to $9,548 (1997-1998)."

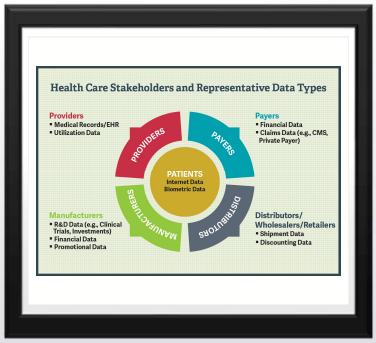
Data is everywhere…

Data is utilized and summarized with statistics frequently in popular media

 For the first time, an influential doctors group is recommending that some children as young as eight be given cholesterol-fighting drugs to ward off future heart problems . . . With one-third of U.S. children overweight and about 17 percent obese, the new recommendations are important,' said Dr. Jennifer Li, a Duke University children's heart specialist."

From cnn.com, Monday July 8th, 2008:

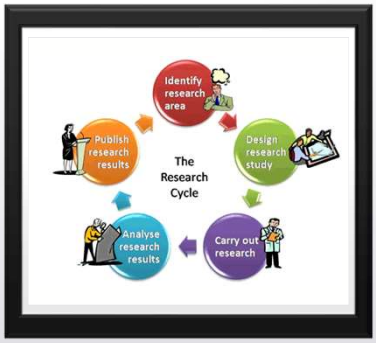
Data is everywhere…



The Good, Bad, and …

- Good data can be analyzed and summarized to provide useful information
- Bad data can be analyzed and summarized to provide incorrect/ harmful/non-informative information

Study design and how data is impacted



Study design and how data is impacted

Data collection

- Data analysis - What statistical methods are appropriate given the data collected? - Dealing with variability (both natural and sampling related):
- Important patterns in data are obscured by variability

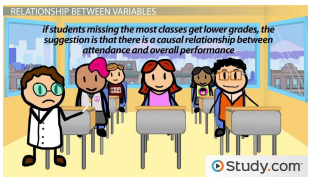
Study design and how data is impacted

- Distinguish real patterns from random variation - Inference: using information from the single study coupled with information about variability to make statement about the larger population/process of interest

Study design and data

- Presentation - What summary measures will best convey the "main messages" in the data about the primary (and secondary) research questions of interest -
- How to convey/ rectify uncertainty in estimates based on the data
- Interpretation - What do the results mean in terms of practice, the program, the population etc?

Examples with important lessons



Example looking at the polio trial

Design: Features of the Polio Trial

- Comparison group
- Randomized
- Placebo controls
- Double blind
- Objective—the groups should be equivalent except for the factor (vaccine) being investigated

Reference:
http://www.stat.luc.edu/StatisticsfortheSciences/MeierPolio.htm

Polio trial example

Such Great Imbalance by Chance?

- Polio cases - Vaccine—82 - Placebo—162
- Statistical methods tell us how to make these probability calculations

Source: Meier, P. (1972), "The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Poliomyelitis Vaccine," In J. Tanur (Editor), Statistics: A Guide to the Unknown. Holden-Day.

Polio trial

Question - There were almost twice as many polio cases in the placebo compared to the vaccine group - Could the results be due to chance?

How We Use Statistics and Graphic Displays

Figure 1.1

- Percentage of each food type in the diet of two hedgehogs. The size of the circle indicates the total amount of food in the stomach.

Figure 1.2

The abundance of terrestrial flatworms in 100 surveyed plots.

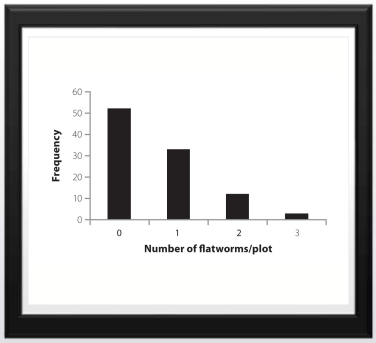


Figure 1.3

- The effect of diet and medication on systolic blood pressure. Data are represented as the mean ± SE.
(Imaginary data)

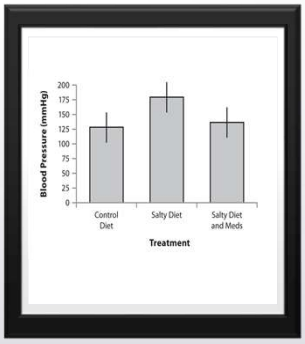


Figure 1.4

- Scatterplots.
- (A) The relationship between verbal ACT scores and quantitative ACT scores (imaginary data).
- (B) Log/log plot with fitted least squared regression line demonstrating the relationship between resting metabolic rate and temperature in teleost fish.
- Data from Clarke and Johnston, 1999. *Journal of Animal Ecology* 68:897.

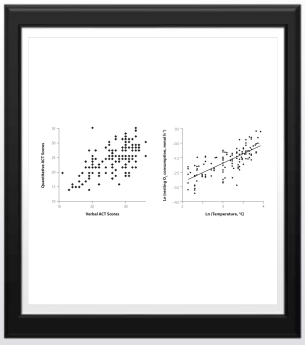
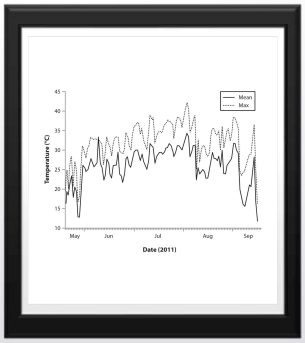


Figure 1.5

- Example of a time-series graph. Air temperatures from 2011 from Springfield, Missouri (daily average and daily maximum). Data from the National Weather Service, graph prepared by K. Reynolds.



Key Terms

bar graph (bar chart)
bias sample
pie chart
population (statistical population)
random sample
sample
scatterplot
statistics
time-series graph

***Chapter 1 and 2 reading**